US009760834B2

(12) United States Patent
Chae et al.

(10) **Patent No.:** **US 9,760,834 B2**
(45) **Date of Patent:** **Sep. 12, 2017**

(54) **DISCOVERY SYSTEMS FOR IDENTIFYING ENTITIES THAT HAVE A TARGET PROPERTY**

(71) Applicant: **Hampton Creek, Inc.**, San Francisco, CA (US)

(72) Inventors: **Lee Chae**, San Francisco, CA (US); **Josh Stephen Tetrick**, San Francisco, CA (US); **Meng Xu**, San Francisco, CA (US); **Matthew D. Schultz**, San Francisco, CA (US); **Chuan Wang**, San Francisco, CA (US); **Nicolas Tilmans**, San Francisco, CA (US); **Michael Brzustowicz**, San Francisco, CA (US)

(73) Assignee: **Hampton Creek, Inc.**, San Francisco, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **15/282,052**

(22) Filed: **Sep. 30, 2016**

(65) **Prior Publication Data**

US 2017/0091637 A1 Mar. 30, 2017

Related U.S. Application Data

(60) Provisional application No. 62/235,174, filed on Sep. 30, 2015, provisional application No. 62/401,787, filed on Sep. 29, 2016.

(51) **Int. Cl.**
G06N 5/04 (2006.01)
G06N 99/00 (2010.01)

(52) **U.S. Cl.**
CPC *G06N 5/048* (2013.01); *G06N 99/005* (2013.01)

(58) **Field of Classification Search**
CPC G06N 5/04
(Continued)

(56) **References Cited**

U.S. PATENT DOCUMENTS

2005/0141769 A1 6/2005 Ho et al.
2007/0275397 A1* 11/2007 Wehrman C12N 15/1055
435/6.12
(Continued)

OTHER PUBLICATIONS

Thomas G. McCloud ("High Throughput Extraction of Plant, Marine and Fungal Specimens for Preservation of Biologically Active Molecules" 2010).*
(Continued)

Primary Examiner — Lut Wong
(74) *Attorney, Agent, or Firm* — Morgan, Lewis & Bockius LLP

(57) **ABSTRACT**

Systems and methods for assaying a test entity for a property, without measuring the property, are provided. Exemplary test entities include proteins, protein mixtures, and protein fragments. Measurements of first features in a respective subset of an N-dimensional space and of second features in a respective subset of an M-dimensional space, is obtained as training data for each reference in a plurality of reference entities. One or more of the second features is a metric for the target property. A subset of first features, or combinations thereof, is identified using feature selection. A model is trained on the subset of first features using the training data. Measurement values for the subset of first features for the test entity are applied to thereby obtaining a model value that is compared to model values obtained using measured values of the subset of first features from reference entities exhibiting the property.

29 Claims, 18 Drawing Sheets